Cross-lingual retrieval system and method that utilizes stored pair data in a vector space model to process queries

Patent number: US6321189

Publication date: 2001-11-20

Inventor: MASUICHI HIROSHI (JP); UMEMOTO HIROSHI (JP); TATENO MASAKAZU (JP)

Applicant: FUJI XEROX CO LTD (US)

Classification:

- international: ***G06F17/28; G06F17/30;*** **G06F17/28; G06F17/30;** (IPC1-7): G06F17/28; G06F17/30

- european: G06F17/27M; G06F17/28D4; G06F17/30H2

Application number: US19990343543 19990630

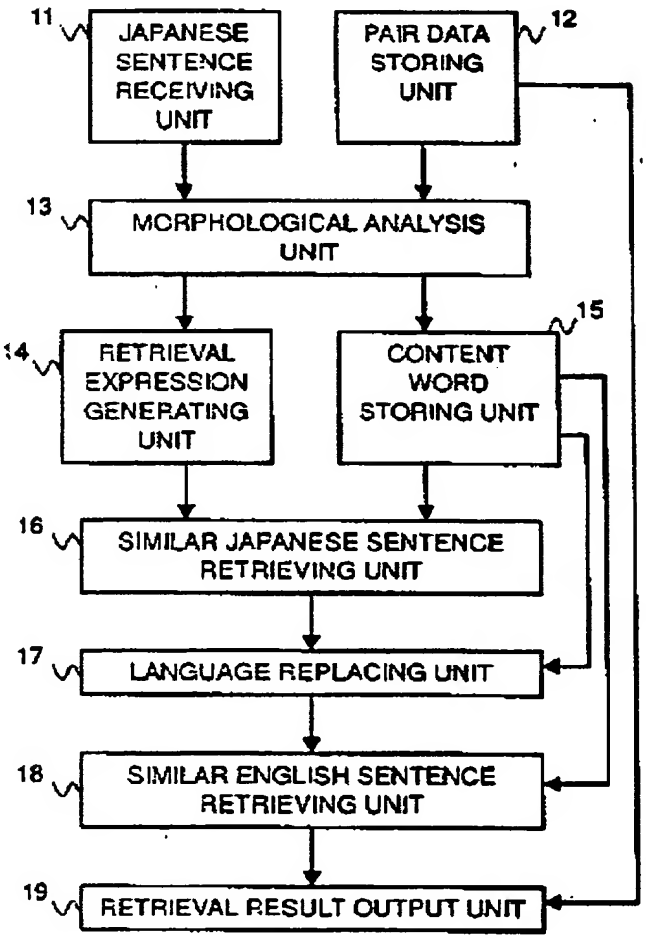
Priority number(s): JP19980202788 19980702

Also published as:

JP2000020524 (A)

Abstract of **US6321189**

The cross-lingual retrieval system of the present invention retrieves second language sentences which are appropriate translations of a query written in a first language using pair data without being greatly influenced by the difference in expression between the query and sentences to be the object of retrieval, length of the query or sentences to be the object of retrieval or the like. In the system, a pair data storing unit stores multiple pairs each having a first language sentence and a second language sentence having the same meaning. As a query receiving unit receives a query written in the first language, a first retrieving unit retrieves first language sentences similar to the query from a set of first language sentences stored in the pair data storing unit. A second retrieving unit then retrieves second language sentences similar to second language sentences paired with the respective first language sentences retrieved by the first retrieving unit from a set of second language sentences stored in the pair data storing unit. That is, a first retrieval of first language sentences from the pair data is performed based on a query, and then a second retrieval of second language sentences from the pair data is performed based on the result of the first retrieval.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号
特開2000－20524
（P2000－20524A）
(43)公開日　平成12年１月21日(2000.1.21)

(51)Int.Cl.'	識別記号	FI		テーマコード'(参考)
G06F 17/28		G06F 15/38	Z	5B075
17/30		15/403	350C	5B091

審査請求　有　　請求項の数22　ＦＤ　（全 18 頁）

(21)出願番号　特願平10－202788

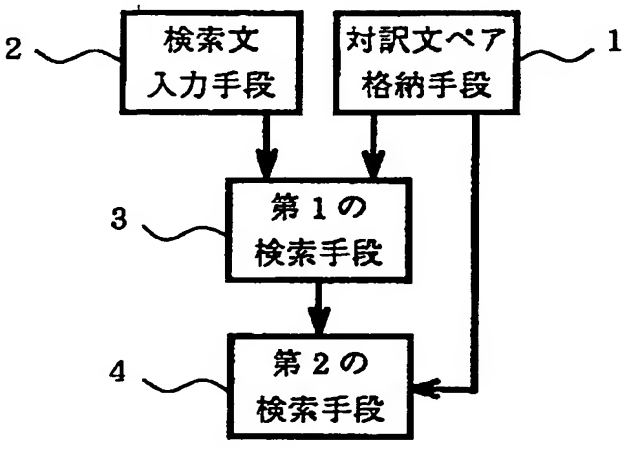
(22)出願日　平成10年７月２日(1998.7.2)

(71)出願人　000005496
富士ゼロックス株式会社
東京都港区赤坂二丁目17番22号
(72)発明者　増市　博
神奈川県足柄上郡中井町境430　グリーンテクなかい　富士ゼロックス株式会社内
(72)発明者　梅基　宏
神奈川県足柄上郡中井町境430　グリーンテクなかい　富士ゼロックス株式会社内
(74)代理人　100098132
弁理士　守山　辰雄

最終頁に続く

(54)【発明の名称】　対訳文検索装置

(57)【要約】
【課題】　表現の相違、文の長さ等に大きく影響されることなく、対訳文ペアを用いて第１言語の検索質問文からより適切な訳文たる第２言語文を検索する。
【解決手段】　対訳文ペア格納手段１に第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納しておき、検索文入力手段２から第１言語で書かれた検索質問文を受け付けると、第１の検索手段３が検索質問文に基づいて対訳文ペア格納手段１に格納されている第１言語で書かれた文の集合を対象として検索処理する。第１の検索手段３により検索された第１言語で書かれた文に対応して対訳文ペア格納手段１に格納されている第２言語で書かれた訳文に類似する文を、第２の検索手段４が対訳文ペア格納手段１に格納されている第２言語で書かれた訳文の集合を対象として検索する。訳文ペアに対する第１言語入力文による検索を行い、この検索結果に対応する第２言語文を用いて、訳文ペアから第２言語文の類似検索を行う。



1

【特許請求の範囲】

【請求項１】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索装置において、

第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、

第１言語で書かれた検索質問文を受け付ける検索文入力手段と、

受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、

第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、当該対訳文ペア格納手段に格納されている第２言語で書かれた訳文の集合を対象として検索する第２の検索手段と、

を有することを特徴とする対訳文検索装置。

【請求項２】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた文を検索する対訳文検索装置において、

第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、

第２言語で書かれた文を複数格納する文格納手段と、

第１言語で書かれた検索質問文を受け付ける検索文入力手段と、

受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、

第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、当該対訳文ペア格納手段に格納されている第２言語で書かれた訳文集合及び文格納手段に格納されている第２言語で書かれた文集合を対象として検索する第２の検索手段と、

を有することを特徴とする対訳文検索装置。

【請求項３】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた文を検索する対訳文検索装置において、

第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、

第２言語で書かれた文を複数格納する文格納手段と、

第１言語で書かれた検索質問文を受け付ける検索文入力手段と、

受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、

第１の検索手段により検索された第１言語で書かれた文

2

に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、文格納手段に格納されている第２言語で書かれた文集合を対象として検索する第２の検索手段と、

を有することを特徴とする対訳文検索装置。

【請求項４】　請求項１乃至請求項３のいずれか１項に記載の対訳文検索装置において、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から所定の基準に基づいた重要語を抽出し、当該重要語を用いて第２言語で書かれた類似文を検索することを特徴とする対訳文検索装置。

【請求項５】　請求項１乃至請求項３のいずれか１項に記載の対訳文検索装置において、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から重要語を抽出するとともに重要語に重要度を付与し、当該重要語及び重要度を用いて第２言語で書かれた類似文を検索し、

更に、対訳文ペア格納手段に格納されている第２言語で書かれた文の集合Ａと、第１の検索手段で検索された第１言語で書かれた文に対応する第２言語で書かれた文の集合Ｂと、集合Ｂ中に出現する全単語の集合Ｃに関して、

集合Ｂに含まれる文の数である第１の値と、集合Ｂ中に出現する単語を重要語候補として各重要語候補を含む集合Ｂ中の文の数である第２の値と、各重要語候補を含む集合Ａ中の文の数である第３の値を求め、これら３種の値を変数として各重要語候補の重要度を算出し、これら重要度に基づいて重要語候補中から重要語が決定されることを特徴とする対訳文検索装置。

【請求項６】　請求項２又は請求項３に記載の対訳文検索装置において、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から重要語を抽出するとともに重要語に重要度を付与し、当該重要語及び重要度を用いて第２言語で書かれた類似文を検索し、

更に、対訳文ペア格納手段に格納されている第２言語で書かれた文の集合と文格納手段に格納されている第２言語で書かれた文の集合の和である集合Ａと、第１の検索手段で検索された第１言語で書かれた文に対応する第２言語で書かれた文の集合Ｂと、集合Ｂ中に出現する全単語の集合Ｃに関して、

集合Ｂに含まれる文の数である第１の値と、集合Ｂ中に出現する単語を重要語候補として各重要語候補を含む集合Ｂ中の文の数である第２の値と、各重要語候補を含む集合Ａ中の文の数である第３の値を求め、これら３種の値を変数として各重要語候補の重要度を算出し、これら重要度に基づいて重要語候補中から重要語が決定される

ことを特徴とする対訳文検索装置。

【請求項７】　請求項５又は請求項６に記載の対訳文検索装置において、

第２の検索手段は、集合Ａ中に含まれる文書の数をＭとし、第１の値をα、重要語候補ごとの第２の値をβ、重要語候補ごとの第３の値をγとした場合に、

拡張相互情報量＝ｌｏｇ｜（Ｍβ）／（$\alpha\gamma$）｜

拡張Ｔ－ｓｃｏｒｅ＝Ｍ｜（Ｍ$\beta-\alpha\gamma$）／（$\alpha\gamma$）｜

拡張Ｄｉｃｅ－ｃｏｅｆｆｉｃｉｅｎｔ＝２β／（$\alpha+\gamma$）

のいずれかの値を各重要語候補の重要度とすることを特徴とする対訳文検索装置。

【請求項８】　請求項１乃至請求項４のいずれか１項に記載の対訳文検索装置において、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた文に基づいて、第２言語で書かれた訳文に類似する文を検索する際に、ベクトル空間モデルを用いることを特徴とする対訳文検索装置。

【請求項９】　請求項１乃至請求項８のいずれか１項に記載の対訳文検索装置において、

第１の検索手段は、受け付けた検索質問文に類似する文を検索するとともに当該類似する文に重要度を付与し、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた文及び第１の検索手段によって付与された重要度に基づいて、第２言語で書かれた訳文に類似する文を検索することを特徴とする対訳文検索装置。

【請求項１０】　請求項１乃至請求項９のいずれか１項に記載の対訳文検索装置において、

検索文入力手段は、第１言語で書かれた検索質問文を受け付けるとともに、当該検索質問文を複数の単語或いは節に分割し、

第１の検索手段は、分割された各単語或いは各節を用いて、受け付けた検索質問文に類似する第１言語で書かれた文を検索し、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を検索し、

更に、各単語或いは各節ごとに第２の検索手段により検索された第２言語で書かれた複数の文の中から、検索結果を所定の重要度を基準として選択する検索結果統合手段を有することを特徴とする対訳文検索装置。

【請求項１１】　請求項１、請求項２、請求項４、請求項５、請求項８、請求項９、請求項１０のいずれか１項に記載の対訳文検索装置において、

第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文と共に、当該第１言語で書かれた文も対訳文ペア格納手段から取得し、当該第１言語で書かれた文と第２言語で書かれた訳文とからなる対訳文ペアに類似する対訳文ペアを対訳文ペア格納手段に格納されている対訳文ペアの集合から検索することを特徴とする対訳文検索装置。

【請求項１２】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、

第１言語で書かれた検索質問文を受け付ける検索文入力機能と、

メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、

第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として検索する第２の検索機能と、

をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。

【請求項１３】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、

第１言語で書かれた検索質問文を受け付ける検索文入力機能と、

メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、

第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合及び当該ペアデータとは別個にメモリに格納されている第２言語で書かれた文の集合を対象として検索する第２の検索機能と、

をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。

【請求項１４】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、

第１言語で書かれた検索質問文を受け付ける検索文入力機能と、

メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用い

て、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、
第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータとは別個にメモリに格納されている第２言語で書かれた文の集合を対象として検索する第２の検索機能と、
をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。

【請求項１５】　請求項１２乃至請求項１４のいずれか１項に記載の対訳文検索プログラムを記憶した記憶媒体において、
記憶媒体にはペアデータが読み出し自在に記憶されており、
対訳文検索プログラムは、当該ペアデータを記憶媒体から読み出してコンピュータに備えられているメモリに格納する機能を含んでいることを特徴とする対訳文検索プログラムを記憶した記憶媒体。

【請求項１６】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、
第１言語で書かれた検索質問文を受け付け、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索し、
第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索することを特徴とする対訳文検索方法。

【請求項１７】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、
第１言語で書かれた検索質問文を受け付け、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索し、
第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合及び当該ペアデータとは別個に用意されている第２言語で書かれた文の集合を対象として第２の検索することを特徴とする対訳文検索方法。

【請求項１８】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、

第１言語で書かれた検索質問文を受け付け、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索し、
第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータとは別個に用意されている第２言語で書かれた文の集合を対象として第２の検索することを特徴とする対訳文検索方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、第１言語と第２言語との２種の言語間の対訳文を検索する装置に関し、特に、第１言語と第２言語との対訳文を用いて、第１言語で書かれた検索質問文により第１言語についての検索を行い、更に、当該検索結果に基づいて第２言語についての類似文検索を行う対訳文検索装置に関する。

【０００２】

【従来の技術】コンピュータ性能の向上、電子辞書の整備、自然言語処理技術の進歩等に伴い、これまで多くの機械翻訳技術の提案がなされてきた。しかしながら、未だ十分な翻訳精度を持つ機械翻訳システムが実現されているとは言い難い状況にある。

【０００３】（従来技術１）そこで、翻訳元の言語（第１言語）と翻訳先の言語（第２言語）の対訳文のペアを多数用意しておき、対訳文ペアの第１言語文から第１言語入力文に類似する文を検索し、この検索結果としての第１言語文に対応する第２言語文を対訳文ペアから出力し、この出力された第２言語文をユーザに参照させることによって、第１言語入力文の翻訳の質を高めようとする手法が提案されている。第１言語入力文と類似する文を対訳文ペアの第１言語文の集合中から得る方法として、共通に用いられている語の多いものを類似度の高い文とする方法や、例えば特開平９－５０４３５号公報に記載されるように、類似文書検索手法の一つであるベクトル空間モデルに基づいた検索手法を用いて、第１言語入力文に対応するベクトルと距離の近いベクトルを持つ第１言語文を類似度の高い文とする方法が提案されている。

【０００４】以下、文献「Ｄａｖｉｄ　Ｅｌｌｉｓ，”情報検索論”，丸善株式会社，ｐｐ．５３－５７（１９９４）」に記載されるベクトル空間モデルにより入力文と類似度の高い文を得る手法について説明する。ベクトル空間モデルにおいては、検索対象となる各文と検索式として与えられる入力文の両者をベクトルとして表現する。検索対象の文がＮ文存在し、Ｎ文中に含まれる全単語がＭ種類（Ｗ1，Ｗ2，・・・，ＷＭ）あるとして、各文（Ｓ1，Ｓ2，・・・，ＳＮ）に対応するベクトルは、それぞれ式１の通りにＭ次元のベクトルとして定義でき

7

る。ただし、Tijは文Si中に単語Wjが存在していれば1、存在していなければ0である。

S1＝（T11，T12，・・・，T1M）、
S2＝（T21，T22，・・・，T2M）、
・・・
SN＝（TN1，TN2，・・・，TNM）、

【０００６】同様に検索入力文Qに対応するベクトルは式２の通りに定義できる。ただし、Tiは入力文Q中に単語Wiが存在していれば1、存在していなければ0である。なお、ここではベクトルの各要素を1，0の2値

Q＝（T1，T2，・・・，TM）

【０００８】ベクトル空間モデルでは、ベクトルQと距離の近い（距離の値が大きい）ベクトルSiに対応する文Siを入力文Qと類似度の高い文とし、検索結果として類似度の高い文から順に出力する。ベクトルQとベクトルSiの距離D（Q，Si）は、式３の計算式に従って行う。ただし、（V，U）はベクトルVとベクトルUの内積である。なお、通常、ベクトル空間モデルでは、計

D（Q，Si）＝（Q，Si）／（（Q，Q）（Si，Si））$^{1/2}$　・・・（式３）

【０００５】
【数１】

・・・（式１）

としているが、単語の各文中での重要度に応じて実数値を割り当てる場合もある。

【０００７】
【数２】

・・・（式２）

算に用いる単語W1，W2，・・・，WMを自立語のみとし、助詞や助動詞等の付属語を考慮しないことが多い。また、自立語であっても、英語のｂｅ動詞のようなありふれた語（ストップワード）は考慮しないことが多い。

【０００９】
【数３】

【００１０】（従来技術２）上記の（従来技術１）と同様の効果を得るために、第１言語の入力文中の各単語を、辞書を用いて第２言語の単語や熟語に機械的に置き換え、この第２言語による単語や熟語の集合を用いて第２言語の文集合中から該当する文を検索し、得られた第２言語文をユーザに参照させることによって翻訳の質を高める手法が研究されている。

【００１１】
【発明が解決しようとする課題】しかしながら、上記した（従来技術１）や（従来技術２）にあっては、次のような課題があった。

【００１２】まず、（従来技術１）は、第１言語の入力文中に存在する単語だけを基に類似する第１言語文を得るものである。このため、第１言語入力文の翻訳文として適切な第２言語文が対訳文ペア集合中に存在している場合であっても、第１言語文の表現が第１言語入力文と異なる場合には、検索結果として適切な第２言語文を得ることができない。この（従来技術１）が効果的であるのは、第１言語入力文中に存在する単語集合とほぼ同一の単語集合を構成要素とする文が対訳文ペア集合中に存在している場合に限られてしまう。このような欠点は、入力文中に含まれる単語の数が少ない程顕著になる。このことからすれば、入力が多くの文から構成される文書のような場合には、対応する文書ベクトルの非零の要素が多くなり（実質的なベクトルの次元が高くなり）、検索結果の信頼性は高くなるといえるが、実際に存在する対訳文データは短い文であることがほとんどであるため、（従来技術１）により適切な対訳文（第２言語文）を得ることは事実上困難である。

【００１３】１つの例として、第１言語を日本語、第２言語を英語とし、「次第に細くなる。」という日本語文を入力文とする場合を考える。この文から自立語を抽出すると、「次第」及び「細い」が得られる。なお、動詞「なる」はストップワードであり、説明から省くことにする。（従来技術１）によれば、上記入力文に類似する日本語文として、「次第」と「細い」の両単語を含む文が得られることになる。しかしながら、入力文に意味的には等しいが、表現の異なる（使用されている単語の異なる）次のような適切な対訳文を得ることはできない。

「先細になる。It tapers down to a point.」　・・・（ａ）、
「先へ行って尖る。It tapers into a sharp point.」　・・・（ｂ）、

【００１４】また、（従来技術２）は、第１言語の入力文中の個々の単語を、辞書を用いて第２言語の単語や熟語に置き換えることにより、参照すべき第２言語文を得るものである。しかしながら、或る第１言語単語を表現することが可能な第２言語の単語や熟語は極めて多様であり、さらに、その中からどの第２言語単語で置き換えるのが適切であるかは第１言語入力文の文意に依存するため、それらを予め決定しておくことは事実上不可能である。したがって、第１言語単語と第２言語単語の対応関係を予め網羅的に辞書の形式で表現することは困難であり、（従来技術２）で適切な対訳文を得ることも困難である。

【００１５】１つの例として、上記の（従来技術１）についての例と同じ条件において、「次第」および「細い」を英語単語あるいは英熟語に置き換えると例えば次のようになる。

「次第」→「gradually, by degrees, little by little, as soon as, order, the state of things, depend on」 ・・・（ｃ）、
「細い」→「thin, narrow, fine, slim, slender」 ・・・（ｄ）、

この（ｃ）および（ｄ）中の英単語を含む文を検索したとしても、上記した（ａ）や（ｂ）に示した適切な英語文を得ることはできない。実際、（ｃ）および（ｄ）中の英単語を中心にして「次第に細くなる。」に対応する英語文を作成すると不自然な文となってしまう。（ａ）や（ｂ）中の英語文に含まれる「taper」は単独で「次第に細くなる。次第に少なくなる。」という意味を持つ単語であるが、（従来技術２）においては、「次第」（あるいは「細い」）の翻訳単語として「taper」が記述されていなければ（ａ）や（ｂ）を得ることはできないことになる。

【００１６】しかしながら、「taper」は、「次第に」と「細くなる」という両者の意味内容が同時に含まれている場合にのみ対応する単語であって、「次第」（あるいは「細い」）単独の翻訳単語としては不適切である。「taper」と同様に単一の語で「次第に」という意味を含む英単語として、「peter（次第に消える）」「vane（次第に弱くなる）」「fade（次第に薄れる）」等様々な単語を挙げることができるが、いずれも同様の理由で「次第」という１単語の翻訳単語としては不適切である。すなわち、ある日本語単語に適切に対応する英単語は入力文の文意に依存し、予め辞書を作成しておくことは不可能であるといえる。

【００１７】なお、前述の特開平９－５０４３５号公報に記載された方法では、ベクトル空間モデルにおいて、各文ではなく、単語ごとに予めベクトルを設定し、文に含まれる各単語に対応するベクトルの総和として文ベクトルを表現している。この場合においても、同一単語が存在しない場合は類似度の値が低くなり、上記した（従来技術１）についての課題は解決されるものではない。さらに、既に述べたように各単語の意味は文意に依存するものであるため、各単語について予め固定的なベクトルを決定しておくことは不可能であり、上記した（従来技術２）と同種の課題を持つことになってしまう。

【００１８】本発明は、上記従来の事情に鑑みなされたものであり、その表現の相違や単語や熟語の数に大きく影響されることなく、対訳文ペアを用いて第１言語の検索質問文からより適切な訳文たる第２言語文を検索することを目的とする。特に、本発明は、第１言語入力文が比較的短い場合であっても、適切な第２言語文を対訳文ペア中から検索することを目的とする。

【００１９】

【課題を解決するための手段】本発明に係る対訳文検索装置では、対訳文ペア格納手段に第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納しておき、検索文入力手段から第１言語で書かれた検索質問文を受け付けると、第１の検索手段が当該検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する。そして、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、第２の検索手段が対訳文ペア格納手段に格納されている第２言語で書かれた訳文の集合を対象として検索する。

【００２０】すなわち、訳文ペアに対する第１言語入力文による検索を行い、この検索結果に対応する第２言語文を用いて、訳文ペアから第２言語文の類似検索を行っている。このように訳文ペアを橋渡しとした第１言語と第２言語との検索を二重に連続して行うことにより、表現の相違や単語や熟語の数に大きく影響されることなく、第１言語の検索質問文からより適切な訳文たる第２言語文を検索することができる。なお、本明細書において、第１の検索処理の対象とする第１言語で書かれた文の集合や、第２の検索処理の対象とする第２言語で書かれた文の集合には、対訳文ペア格納手段（文格納手段）に格納された状態の文データである場合に限らず、これら文データから抽出された節や単語或いは文集合であって元の文データに対応付けられていることにより、検索処理上、対訳文ペア格納手段（文格納手段）に格納された文集合と実質的に同一に扱うことができるデータ集合をも包含している。

【００２１】また、本発明に係る対訳文検索装置では、対訳文ペア格納手段に第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納し、また、文格納手段に第２言語で書かれた文を複数格納しておき、検索文入力手段から第１言語で書かれた検索質問文を受け付けると、第１の検索手段が当該検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する。そして、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、第２の検索手段が対訳文ペア格納手段及び文格納手段に格納されている第２言語で書かれた文の集合を対象として検索する。

【００２２】すなわち、訳文ペアに対する第１言語入力文による検索を行い、この検索結果に対応する第２言語文を用いて、訳文ペア及び別個な第２言語文の集合から第２言語文の類似検索を行っている。この態様においても訳文ペアを橋渡しとした第１言語と第２言語との検索を二重に連続して行っており、これにより、表現の相違や単語や熟語の数に大きく影響されることなく、第１言語の検索質問文からより適切な訳文たる第２言語文を検索することができる。なお、この態様によれば、第２の検索手段による類似文検索に、訳文ペアの他に、例えば

外部の文書集合データベースをもネットワーク等を通して利用し、より多くのデータを用いてより適切なる第２言語の訳文を得ることができる。

【００２３】また、本発明に係る対訳文検索装置では、対訳文ペア格納手段に第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納し、また、文格納手段に第２言語で書かれた文を複数格納しておき、検索文入力手段から第１言語で書かれた検索質問文を受け付けると、第１の検索手段が当該検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する。そして、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、第２の検索手段が文格納手段に格納されている第２言語で書かれた文の集合を対象として検索する。

【００２４】すなわち、訳文ペアに対する第１言語入力文による検索を行い、この検索結果に対応する第２言語文を用いて、訳文ペアとは別個な第２言語文の集合から第２言語文の類似検索を行っている。この態様においても訳文ペアを橋渡しとした第１言語と第２言語との検索を二重に連続して行っており、これにより、表現の相違や単語や熟語の数に大きく影響されることなく、第１言語の検索質問文からより適切な訳文たる第２言語文を検索することができる。なお、この態様によれば、第２の検索手段による類似文検索に、例えば外部の文書集合データベースをネットワーク等を通して利用し、種々なデータを用いてより適切なる第２言語の訳文を得ることができる。

【００２５】なお、本発明に係る対訳文検索装置は、後述するように、第２の検索手段による検索において重要語やその重要度を用いる、第２の検索手段による検索においてベクトル空間モデルを用いる、第２の検索手段により第１言語で書かれた文と第２言語で書かれた訳文とからなる対訳文ペアに類似する対訳文ペアを検索する、第１言語で書かれた検索質問文を複数の単語或いは節に分割して、これら各単語或いは各節ごとに類似文検索を行う、等と言った種々な態様で実現することができる。また、本発明に係る対訳文検索方法は、上記のような訳文ペアを用いた第１言語による第１の検索と、この検索結果に基づいて、訳文ペア及び／又は別個な第２言語文集合を用いた第２言語による第２の検索とを行うことによって、第１言語の検索質問文からより適切な訳文たる第２言語文を検索することができる。

【００２６】また、本発明は、上記した対訳文検索をコンピュータに実行させるためのプログラムを記憶した記憶媒体としても実施でき、第１言語で書かれた検索質問文を受け付ける検索文入力機能と、メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合（及び／又は、当該ペアデータとは別個にメモリに格納されている第２言語で書かれた文の集合）を対象として検索する第２の検索機能と、をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能にＣＤＲＯＭ等の記憶媒体に記憶した。

【００２７】ここで、本発明の記憶媒体において、コンピュータに設けられているメモリや外部のデータベースに予め用意したペアデータを用いるようにしてもよいが、ペアデータを上記の対訳文検索機能プログラムと共に記憶媒体に記憶したパッケージとしてもよい。なお、この場合には、記憶媒体に記憶されたペアデータをコンピュータに利用させるために、当該記憶媒体に記憶される対訳文検索機能プログラムには、ペアデータを記憶媒体から読み出してコンピュータに備えられているメモリに格納する機能が含まれている。

【００２８】より具体的に説明すると、本発明を実施した場合の対訳文検索装置の典型的な態様は図１に示すようなものとなる。すなわち、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段１と、第１言語で書かれた検索質問文を受け付ける検索文入力手段２と、検索質問文に類似する文を対訳文ペア格納手段１に格納されている第１言語で書かれた文の集合から検索する第１の検索手段３と、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段１に格納されている第２言語で書かれた文を入力として、当該文に類似する文を対訳文ペア格納手段１に格納されている第２言語で書かれた文の集合から検索する第２の検索手段４と、を備えて構成される。

【００２９】第１の検索手段３による類似文検索は、例えば、第１言語の検索質問文から自立語を抽出した後、（１）得られた自立語集合を基にベクトル空間モデルに従って得られる類似文（第１言語文）の内から距離の値が所定の閾値よりも大きい文を検索結果とする、又は、（２）後述する拡張相互情報量に基づいた計算によって得られる類似文（第１言語文）の内から拡張相互情報量の合計値が所定の閾値よりも大きい文を検索結果とする、のいずれかの方法で行う。そして、第１の検索手段３から得られた第１言語文の集合中の各第１言語文は、対訳文ペア格納手段１に格納されている対応する第２言語文に置き換えられて、第２の検索手段４に入力される。

【００３０】第２の検索手段４による類似文検索は、例えば、このようにして得られた第２言語文集合から自立語を抽出し、（１）得られた自立語集合を基にベクトル

空間モデルに従って得られる類似文（第２言語文）の内から距離の値が所定の閾値よりも大きい文を検索結果とする、又は、（２）後述する拡張相互情報量に基づいた計算によって得られる類似文（第２言語文）の内から拡張相互情報量の合計値が所定の閾値よりも大きい文を検索結果とする、のいずれかの方法で行う。なお、複数の文を入力として、ベクトル空間モデルを利用する場合には、入力された各文に対応する文ベクトルの総和ベクトルを入力文ベクトルとみなして、単一文入力の場合と同様の計算を行えばよい。

【００３１】第１の検索手段３で行う検索は、入力された検索質問文と類似性のある第１言語文を広範に得ることを目的としている。ベクトル空間モデルによる類似文検索も、後述する拡張相互情報量に基づいた計算による類似文検索も、統計的な手法に基づくものであり、検索質問文中の異なり語数が多いほど結果の信頼性は高くなる。したがって、第１の検索手段３で行う類似文検索は、第２の検索手段４で行う類似文検索に対する入力文の量を増加させ、より漏れの少ない検索とすることを狙いとしている。また、第２の検索手段４で行う検索は、第２言語で類似検索を行うものであり、これにより、第１言語の表現の違いに依存しない第２言語の類似文を得ることが可能となる。

【００３２】このような第１の検索手段３と第２の検索手段４の組み合わせにより、前記した（従来技術１）および（従来技術２）における課題を解決することができる。（従来技術１）における課題は、第１言語入力文と意味的には等しいが表現の異なる第１言語文が対訳文ペア集合中に存在する場合でも、その文を検索できないというものであった。本発明によると、第２言語文集合に対する第２の検索手段４による類似文検索の結果が最終結果となるため、第１言語入力文に含まれる単語を全く含まないものであっても、第１言語入力文と類似性の高い対訳文ペアが存在すれば、それを検索結果として得ることができる。また、（従来技術２）における課題は、或る第１言語単語に対応する第２言語単語を網羅的に記述する辞書を予め作成しておくことが不可能であるというものであった。本発明によると、第１の検索手段３によって第１言語入力文と類似度の高い第１言語文集合を取得し、それらに対応する第２言語文集合を基に第２の検索手段４で類似文検索を行っており、第１言語単語と第２言語単語の対応関係は、第１言語入力文に応じて類似文検索された広範な対訳文ペアから得られる単語情報によって動的に決定される。すなわち、第１言語単語に対応する第２言語単語を辞書として作成しておくことなしに、網羅的な第１言語単語と第２言語単語の対応関係を得ることが可能となる。

【００３３】１つの例として、上記した（従来技術１）についての例と同じ条件で、本発明の第１の検索手段３から、「イギリスの援助が次第に減じた。（British aid tapered off.）」、「フランスワインの購買が次第に先細りになった。（The purchase of French wine tapered off.）」、「彼の筋肉たくましい脚は下に向かってだんだん細くなり、足首はほっそりとしていた。（His muscular legs tapered to slender ankles.）」、「彼の言っていることがはっきりわかってくると拍手がっさいが心もとなげに次第に小さくなった。（When they realized what hewas saying, the applause tapered off uncertainly.）」、と言った日本文が得られる。これらのいずれの文も「次第」あるいは「細い」のいずれかが含まれているため、入力日本語文「次第に細くなる」の類似文として得られるものである。第２の検索手段４は、第１の検索手段３から得られた日本語文集合の各々に対応する英語文の集合を入力として類似文検索を行うが、上記の例文中の英語訳には全て「taper」が含まれているため、第２の検索手段４の検索結果では「taper」を含む文の類似度が高くなり、上記した（ａ）（ｂ）が検索結果として得られることになる。

【００３４】

【発明の実施の形態】本発明をその一実施形態に基づいて具体的に説明する。本実施形態では、本来単語間の類似度として用いる統計量である相互情報量、Ｄｉｃｅ－ｃｏｅｆｆｉｃｅｎｔおよびｔ－ｓｃｏｒｅを拡張することによって、検索式と単語の間の類似度計算を実現している。なお、相互情報量、Ｄｉｃｅ－ｃｏｅｆｆｉｃｅｎｔおよびｔ－ｓｃｏｒｅを単語間の類似度計算に用いた例として、「春野，山崎：辞書と統計を用いた対訳アライメント，情報処理学会自然言語処理研究会研究報告，９６－ＮＬ－１１２，ｐｐ．２３－３０（１９９６）」、「大森，堤，中西：統計情報を用いた対訳単語辞書の作成，言語処理学会第２回年次大会発表論文集，ｐｐ．４９－５２（１９９６）」等を挙げることができる。

【００３５】単語ｗｏｒｄ１と単語ｗｏｒｄ２の間の相互情報量（ＭＩ）は、式４によって定義される。ただし、全検索対象文書数をＭ、ｗｏｒｄ１とｗｏｒｄ２を共に含む文書数をａ、ｗｏｒｄ１のみを含む文書数をｂ、ｗｏｒｄ２のみを含む文書数をｃとした場合、それぞれ出現確率は式５である。

【００３６】

【数４】

$$MI(word1, word2) = \log_2 |prob(word1, word2) / (prob(word1)\ prob(word2))| \quad \cdots (式4)$$

【００３７】

【数５】

$$prob(word1, word2) = a / M$$

prob（word1）＝（a＋b）／M
prob（word2）＝（a＋c）／M　　・・・（式5）

【００３８】これに対して本実施形態では、検索式Ｓと単語ｗｏｒｄの間の相互情報量（ＭＩ’）を、式６によって定義している。ただし、全検索対象文書数をＭ、ｗｏｒｄを含み且つ検索式Ｓから得られる文書の数をａ’、検索式Ｓから得られる文書の内のｗｏｒｄを含まない文書の数をｂ’、ｗｏｒｄを含む文書のうち検索式Ｓから得られる文書を除いた文書の数をｃ’とした場合、それぞれ出現確率は式７である。

【００３９】

【数６】

$$MI'(S, word) = \log_2 |prob(S, word) / (prob(S)\ prob(word))|$$

・・・（式６）

【００４０】

【数７】

prob（S，word）＝a’／M
prob（S）＝（a’＋b’）／M
prob（word）＝（a’＋c’）／M　　・・・（式７）

【００４１】また、相互情報量と同様に単語間の類似度を求める統計量として、Ｄｉｃｅ－ｃｏｅｆｆｉｃｅｎｔおよびｔ－ｓｃｏｒｅを挙げることができる。Ｄｉｃｅ－ｃｏｅｆｆｉｃｅｎｔ（ＤＣ）は式８、ｔ－ｓｃｏｒｅ（ＴＳ）は式９で定義される。

【００４２】

【数８】

DC（word1，word2）＝2prob（word1，word2）／（prob（word1）＋prob（word2））　・・・（式８）

【００４３】

【数９】

TS（word1，word2）＝ M（prob（word1，word2）－prob（word1）prob（word2））／（prob（word1）prob（word2））　・・・（式９）

【００４４】これらについても、相互情報量と同様に、検索式と単語の間の類似度計算するために式１０および式１１に示す拡張を施している。なお、ＭＩ’（Ｓ，ｗｏｒｄ）、ＤＣ’（Ｓ，ｗｏｒｄ）、ＴＳ’（Ｓ，ｗｏｒｄ）のいずれも、その値が大きいほど検索式Ｓと単語ｗｏｒｄの間に高い類似性があることを意味する。

【００４５】

【数１０】

DC’（S，word）＝2prob（S，word）／（prob（S）＋prob（word））
・・・（式１０）

【００４６】

【数１１】

TS’（S，word）＝M（prob（S，word）－prob（S）prob（word））／（prob（S）prob（word））・・・（式１１）

【００４７】本実施形態では、上記の検索式と単語の間の拡張相互情報量（ＭＩ’）、又は、拡張Ｄｉｃｅ－ｃｏｅｆｆｉｃｅｎｔ（ＤＣ’）、又は、拡張ｔ－ｓｃｏｒｅ（ＴＳ’）のいずれかによって、検索式Ｓと単語ｗｏｒｄ（すなわち、これら値の重要度を持った重要語候補）の間の類似度を求めている。具体的には、本実施例では、拡張相互情報量（ＭＩ’）を用いた以下のアルゴリズム［Ｓ０１］～［Ｓ０４］を本実施形態の対訳文検索装置で実行することによって、文集合Ｄを検索対象として、検索式Ｓの類似文検索を行っている。

【００４８】［Ｓ０１］：検索式Ｓで文書集合Ｄを検索し、得られた文集合中に存在する全ての自立語を形態素解析処理（文を単語に分割する処理）を施すことにより抽出する。なお、得られた自立語集合をＷ＝（w1，w2，・・・，wn）とする。

［Ｓ０２］：（式６）により、検索式Ｓと自立語集合Ｗ中の各自立語との間の拡張相互情報量（ＭＩ’（Ｓ，w1），ＭＩ’（Ｓ，w2），…，ＭＩ’（Ｓ，Wn））を求める。

［Ｓ０３］：文集合Ｄ中の全ての文を対象として、自立語集合Ｗの要素wiを含む文に対してＭＩ’（Ｓ，wi）の値を加える計算を、１≦ｉ≦ｎを満たすｉについて繰り返す。

［Ｓ０４］：ＭＩ’（Ｓ，wi）の合計値の高い文から順に出力し、検索式Ｓの類似度文検索の結果とする。

【００４９】それでは、図２を参照して本実施形態に係る対訳文検索装置の構成を説明する。なお、本実施形態では、ＣＤＲＯＭ等の携帯可能な記憶媒体に格納された対訳文検索プログラムをコンピュータの読み取り装置で読み取らせ、当該プログラムをコンピュータに実行させ

ることにより対訳文検索装置を実現しているが、本発明に係る対訳文検索装置は以下に説明する各機能を実現する専用の装置として実施してもよい。また、本実施形態では、第１言語を日本語、第２言語を英語として説明しているが、形態素解析処理が適用可能な言語であればどうような言語であっても同様の効果を得ることができる。

【００５０】日本語文入力手段１１は、類似する英語文を得るためにユーザが入力する日本語で書かれた検索質問文を受け付けるユーザインターフェースを有したプログラムモジュールである。なお、日本語文入力手段１１の他の態様としては、通信回線を介して遠隔地のユーザから検索質問文を受け付ける通信インタフェースを備えてもよい。対訳文ペア格納手段１２は、日本語文と英語文の対訳文のペアを複数格納するメモリを有したプログラムモジュールであり、本例では、これらペアデータは予めコンピュータのメモリに用意しておくが、対訳検索プログラムと共にペアデータを記憶媒体に格納しておき、当該ペアデータをコンピュータのメモリに書き込む或いはコンピュータが当該記憶媒体にアクセスして必要なペアデータを用いるようにしてもよい。なお、本例では、日本語文とそれに対応する英語文とからなる各ペアデータには、それらを一意に特定できる識別子（対訳文ペア識別子）が割り振られている。

【００５１】形態素解析手段１３は、対訳文ペア格納手段１２に格納されている全ての文と、日本語文入力手段１１から受け付けた日本語検索質問文に対して形態素解析処理を行うプログラムモジュールである。形態素解析手段１３は、対訳文ペア格納手段１２に格納されている文の解析結果は対訳文ペア自立語格納手段１５に格納し、日本語文入力手段１１から受け付けた日本語検索質問文の解析結果は検索式作成手段１４に引き渡す。検索式作成手段１４は、日本語検索質問文の形態素解析結果を形態素解析手段１３から受け取って該解析結果から自立語（ただしストップワードは除く）を抽出し、得られた自立語を論理和演算子ＯＲで結合して検索式とするプログラムモジュールである。

【００５２】対訳文ペア自立語格納手段１５は、対訳文ペア格納手段１２に格納されている全ての文の形態素解析結果を形態素解析手段１３から受け取って、該解析結果から自立語（ただしストップワードは除く）を抽出した上で、各対訳文ペア識別子ごとに格納するメモリを有したプログラムモジュールである。日本語類似文検索手段１６は、検索式作成手段１４によって作成された検索式を入力として、対訳文ペア自立語格納手段１５に格納されている日本語単語情報に基づいて、該検索式に類似する複数の日本語文（対訳文ペア識別子）を得るプログラムモジュールである。なお、この類似文を得るための検索に拡張相互情報量が用いられる。

【００５３】言語変換手段１７は、日本語類似文検索手段６によって得られた対訳文ペア識別子に対応する英単語の全てを対訳文ペア自立語格納手段１５から取得して、それらを論理和演算子ＯＲで結合して英単語による検索式を作成するプログラムモジュールである。英語類似文検索手段１８は、言語変換手段１７によって作成された検索式を入力として、対訳文ペア自立語格納手段１５に格納されている英語単語情報に基づいて、該検索式に類似する複数の英語文（対訳文ペア識別子）を得るプログラムモジュールである。なお、この類似文を得るための検索に拡張相互情報量が用いられる。検索結果出力手段１９は、英語類似文検索手段１８の検索結果を受け取って、それらをユーザに対して表示するユーザインターフェースを有したプログラムモジュールである。

【００５４】図３には、対訳文ペア格納手段１２に複数格納されている対訳文ペア（ペアデータ）の一例を示してあり、各対訳文ペアは対訳文ペア識別子で一意に特定される日本語文とそれに対応する英語文とから成っている。図４には、対訳文ペア自立語格納手段１５に格納される対訳文ペアの形態素解析結果の一例を示してあり、対訳文ペアの日本語文から抽出された日本語自立語とそれに対応する英語文から抽出された英語自立語とが、元の対訳文ペアと同一の対訳文ペア識別子で特定されている。すなわち、対訳文ペア自立語格納手段１５に格納される対訳文ペアと対訳文ペア格納手段１２に格納されている対訳文ペアとは、対訳文ペア識別子によって一意に対応付けられている。なお、本実施形態では、類似文検索を行う前に、対訳文ペア自立語格納手段１５の格納内容を得るために、対訳文ペア格納手段１２に格納されている全ての対訳文ペアに対して形態素解析処理を施しておく。

【００５５】図５には、上記構成の対訳文検索装置によって実行されるアルゴリズムをフローチャートで示してあり、当該アルゴリズムを実行することによって、日本語文入力手段１１に入力された日本語検索質問文に類似する対訳英語文が得られる。まず、日本語文入力手段１１が入力された日本語入力文（検索質問文）Ｑを受け付けると（ステップＳ１）、形態素解析手段１３が日本語検索質問文Ｑに形態素解析処理を施して、単語に分割する（ステップＳ２）。そして、検索式作成手段１４が、日本語検索質問文Ｑから得られた単語の内からストップワード以外の自立語を抽出し、各自立語を論理和演算子ＯＲで結合して検索式Ｓとする（ステップＳ３）。

【００５６】次いで、日本語類似文検索手段１６が、検索式Ｓを入力として、対訳文ペア格納手段１２に格納されている対訳文ペアの日本語文を対象に通常の検索（例えば、キーワードマッチング）を行い、検索式Ｓ中のいずれかの単語を含む日本語文を検索して得られた検索結果数をＭとする（ステップＳ４）。そして、日本語類似文検索手段１６が当該Ｍが０であるか否かを判断し（ステップＳ５）、Ｍが０である場合には、入力された日本

語検索質問文Ｑに類似する英語文は検索対象のデータ中に存在しない旨を検索結果出力手段１９から表示出力して処理を終了する（ステップＳ１４）。なお、検索対象の対訳文ペア中に入力日本語文Ｑ中の単語すら含んでいない場合には、当該対訳文ペア中に類似文が存在する可能性はほとんど無いと言えるので、本実施形態では、このような通常の検索（ステップＳ４）を前処理的に行うことによって以後の類似文検索を無駄に行わないようにしている。

【００５７】一方、Ｍが０でない場合には、日本語類似文検索手段１６が、検索式Ｓを入力として、対訳文ペアの日本語文を対象に類似文検索を行う（ステップＳ６）。すなわち、上記したアルゴリズム［Ｓ０１］～［Ｓ０４］を実行し、対訳文ペアの日本語文の集合を文書集合Ｄとし、対訳文ペア自立語格納手段１５に格納されている日本語の自立語集合を自立語集合Ｗとして、類似文検索を行う。そして、日本語類似文検索手段１６が、この検索結果の内で閾値Ｔ（予め設定した非負の定数）を越える拡張相互情報量の合計値を持つ対訳文ペア識別子の集合をＥとし（ステップＳ７）、この識別子集合Ｅの要素数が０であるか否かを判断する（ステップＳ８）。

【００５８】この結果、識別子集合Ｅの要素数が０である場合には、適切な類似文が得られないのでステップＳ１４へ進んで処理を終了する一方、識別子集合Ｅの要素数が０でない場合には、言語変換手段１７が、識別子集合Ｅ中の各識別子に対応する全ての英単語を対訳ペア自立語格納手段１５から抽出し、論理和演算子ＯＲで結合して検索式Ｓ’とする（ステップＳ９）。そして、英語類似文検索手段１８が、検索式Ｓ’を入力として、対訳文ペアの英語文を対象として類似文検索を行う（ステップＳ１０）。すなわち、上記したアルゴリズム［Ｓ０１］～［Ｓ０４］を実行し、対訳文ペアの英語文の集合を文書集合Ｄとし、対訳文ペア自立語格納手段１５に格納されている英語の自立語集合を自立語集合Ｗとして、類似文検索を行う。

【００５９】そして、英語類似文検索手段１８が、この検索結果の内で閾値Ｔ’（予め設定した非負の定数）を越える拡張相互情報量の合計値を持つ対訳文ペア識別子の集合をＥ’とし（ステップＳ１１）、この識別子集合Ｅ’の要素数が０であるか否かを判断する（ステップＳ１２）。この結果、識別子集合Ｅ’の要素数が０である場合には、適切な類似文が得られないのでステップＳ１４へ進んで処理を終了する一方、識別子集合Ｅ’の要素数が０でない場合には、結果出力手段１９が、識別子集合Ｅ’に対応する英語文（あるいは対訳文ペア）を、拡張相互情報量の合計値の大きいものから順に日本語入力文Ｑの類似文として表示出力して、処理を終了する（ステップＳ１３）。

【００６０】なお、上記した実施形態では、英語類似文検索手段１６（第２の検索手段）の検索対象を対訳文ペア中の英語文としているが、第２の検索手段の検索対象は対訳文ペアである必要はなく、対応する日本語文を持たない英語文の集合、或いは、対訳文ペア中の英語文集合と対応する日本語文を持たない英語文の集合との両方としてもよい。この場合には、対訳文ペア格納手段１２とは別に、英語文のみを複数格納する英語文格納手段を設け、第２の検索手段の検索対象を、英語文格納手段中の英語文のみ、或いは対訳文ペア格納手段１２中の英語文および英語文格納手段中の英語文とすればよい。なお、このようにした場合には、上記したステップＳ１０で行う類似文検索アルゴリズム［Ｓ０１］～［Ｓ０４］の処理において、文書集合Ｄを「英語文格納手段中の英語文」或いは「対訳文ペア格納手段１２中の英語文および英語文格納手段中の英語文」とし、自立語集合Ｗを「英語文格納手段中の英語文から抽出した全ての自立語」或いは「対訳文ペア格納手段１２中の英語文および英語文格納手段中の英語文から抽出した全ての自立語」とすればよい。

【００６１】また、英語類似文検索手段１８が、第２の検索手段として、英語文のみならず英語文と日本語文の両者を対象として類似文検索を行うようにしてもよい。この場合には、上記のステップＳ９の処理において、言語変換手段１７が各識別子に対応する英単語および日本語単語の両者を抽出して検索式Ｓ’を作成し、ステップＳ１０の処理において、英語文とそれに対応する（同一の識別子を持つ）日本語文の両者をまとめて単一の文とみなし、それら対訳文ペアの集合をアルゴリズム［Ｓ０１］～［Ｓ０４］中の文書集合Ｄとし、検索対象の英語文および日本語文から抽出した全ての自立語を自立語集合Ｗとして処理を行えばよい。

【００６２】また、上記の実施形態では、拡張相互情報量に基づいた類似文検索を行ったが、拡張Ｔ－ｓｃｏｒｅ、拡張Ｄｉｃｅ－ｃｏｅｆｆｉｃｉｅｎｔ、更には、ベクトル空間モデルによる類似文検索を行う場合でも同様の効果が得られる。また、上記の実施形態では、日本語類似文検索手段１６（第１の検索手段）と英語類似文検索手段１８（第２の検索手段）とが等しいアルゴリズムで類似文検索を行ったが、第１の検索手段は、第１言語の検索質問文と関連する文を広範に得ることが目的であるので、本発明では、例えば、上記のステップＳ４における検索結果を直接日本語類似文検索手段１６の検索結果として、英語類似文検索手段１８が用いるようにしてもよく、同様の効果が得られる。すなわち、本発明では、第１の検索手段による検索処理には、類似文検索と通常の検索とのいずれを採用してもよい。

【００６３】また、上記の実施形態では、日本語類似文検索手段１６の検索結果として得られる対訳文ペア識別子に与えられる拡張相互情報量（重要度）の合計値（識別子の拡張相互情報量値）を英語類似文検索手段１８で

の類似文検索処理に利用していないが、日本語類似文検索手段１６で得られる拡張相互情報量を英語類似文検索手段１８での類似文検索処理に利用して、日本語入力文との類似度をより正確に反映した英語対訳文を得ることもできる。具体的には、この重要度を利用するためには、上記のステップＳ１０の処理におけるアルゴリズム［Ｓ０１］～［Ｓ０４］で式（６）の計算を行う際に、式（７）中のａ’、ｂ’、ｃ’を文書数とする代わりに、日本語類似文検索手段１６で得られた各文識別子の拡張相互情報量値の合計値とすればよい。

【００６４】また、上記の実施形態では、日本語の入力文から英語類似対訳文を得るものであるが、本発明では、構文解析処理により、日本語入力文を意味的に独立な複数の節や単語に分割し、それぞれの節や単語から英語類似対訳文を求めて、日本語入力文の意味内容をより正確に反映した検索を行うこともできる。この場合には、各節や各単語に対して上記のステップＳ１～Ｓ１４の処理を行い、それぞれの処理において英語類似文検索手段１８から得られる各識別子の拡張相互情報量値の合計を最終的値として、その値の大きいものから順に検索結果として出力するようにすればよい。

【００６５】図６には、日英の対訳文３６万ペアを用いて、上記した実施形態の対訳文検索装置により「雲が立ち込めてきた。」という日本語入力文の英語対訳文を検索した結果を示してある。なお、同図には、得られた検索結果の上位２０文から「雲が立ち込めてきた。」の英訳を行う際の参照英語文として適切であると判断できたものを抜き出して示してある。この例に示すように、対訳として適切な英語文を７文得ることができた。なお、各文に付加した数字は、それぞれの文が検索結果の上位何番目に位置していたかを示すものである。例えば、「Clouds blanketed the sky.」という文は、３６万文の検索対象の中から９番目に類似度の高い文であるとして検索されたことを示している。

【００６６】「雲が立ち込めてきた。」を形態素解析することにより、自立語「雲」と「立ち込める」を得ることができるが、本例のように「雲」と「立ち込める」を同時に含む日本語文が３６万文の中に存在しない場合には、（従来技術１）によっては適切な検索結果を得ることはできない。実際、図６に示す日本語文には「立ち込める」という表現は存在しておらず、これらの適切な対訳文を（従来技術１）によって得ることはできない。また、「雲」に対応する英単語は「cloud」であり、「立ち込める」に対応する英単語／英熟語は「hang over, envelop, shroud, screen」であるが、本例のように「cloud」と「hang over, envelop, shroud, screen」のいずれかを同時に含む英語文が３６万文の中に存在しない場合には、（従来技術２）によっては適切な検索結果を得ることはできない。実際、図６に示す英語文には「hang over, envelop, shroud, screen」のいずれの表現も存在しておらず、これらの適切な対訳文を（従来技術２）によって得ることはできない。

【００６７】図７には、同様に、「本名を伏せておくことにした。」という日本語入力文の英語対訳文検索を行った結果を示してある。この例に示すように、対訳として適切な英語文を５文得ることができた。本例のように「本名を伏せておくことにした。」から得られる自立語「本名」および「伏せる」を共通に含む日本語文が検索対象の文集合中に存在しない場合には、「本名」および「伏せる」に対応する英単語／英熟語である「real name, autonyl」のいずれかと「hide, conceal, keep secret」のいずれかを共通に含む英語文も存在しない。したがって、（従来技術１）および（従来技術２）によって適切な英語文を検索することはできない。

【００６８】図８には、上記と同様の条件で、本発明による対訳文検索と（従来技術１）による対訳文検索を実行した結果を示してある。同図の左欄に示す２０文の日本語入力文でそれぞれ検索を行い、それぞれ上位２０位中に適切な検索結果が何文含まれているかを示してあり、本発明の対訳文検索によれば、（従来技術１）の対訳文検索と比較して、適切な対訳文を３倍以上多く検索できた。以上の実験結果からも明らかなように、本発明によれば、従来の技術では検索結果として得ることができなかった適切な類似対訳文をユーザに提示することができる。

【００６９】

【発明の効果】以上説明したように、本発明によると、第１言語文とそれに対応する第２言語文との対訳ペアを用いて、第１言語文に対する第１の検索と第２言語文に対する第２の検索とを組み合わせて実行するようにしたため、（１）比較的短い第１言語の検索質問文からでも、広範な対訳文ペア情報から検索漏れの少ない対訳文検索を行うことができ、（２）第１言語の検索質問文の表現の差異に依存することなしに、適切な対訳文検索を行うことができ、（３）予め作成された辞書を必要とせず、広範な対訳文ペア情報から第１言語単語と第２言語単語の対応関係を動的に取得することができるため、第１言語の検索質問文の文意に応じた対訳文検索を行うことができるといった効果を得ることができる。

【図面の簡単な説明】

【図１】　本発明に係る典型的な対訳文検索装置の構成を示す図である。

【図２】　本発明の一実施形態に係る対訳文検索装置の構成を示す図である。

【図３】　対訳文ペアの一例を示す図である。

【図４】　対訳文ペアを形態素解析した結果の一例を示す図である。

【図５】　本発明の一実施形態に係る検索処理手順を示すフォローチャートである。

【図６】　本発明の実施例に係る対訳文検索の結果を示

す図である。

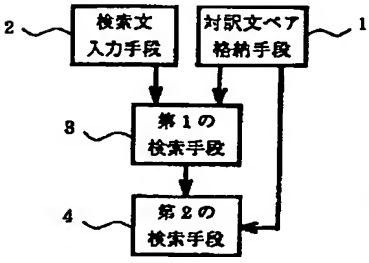
【図７】　本発明の実施例に係る対訳文検索の結果を示す図である。

【図８】　本発明の実施例と従来技術とによる対訳文検索の結果を示す図である。

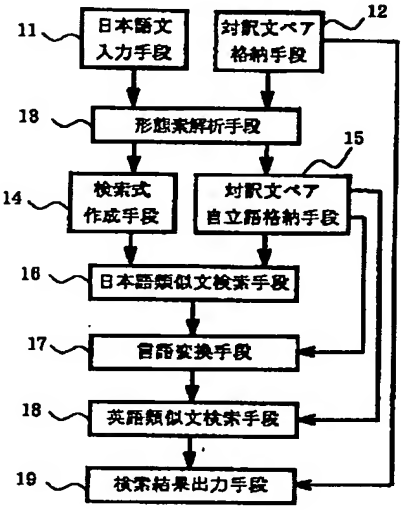
【符号の説明】

１１・・・日本語文入力手段、　１２・・・対訳文ペア格納手段、１３・・・形態素解析手段、　１４・・・検索式作成手段、１５・・・対訳文ペア自立語格納手段、　１６・・・日本語類似文検索手段、１７・・・言語変換手段、　１８・・・英語類似文検索手段、１９・・・検索結果出力手段、

【図１】



【図２】



【図３】

識別子	日本語	英語
1	計算はそろばんで玉を動かして行なわれる。	Computations are performed on the abacus by manipulating the counters on it.
2	私はどうでもいいという気持ちに取りつかれた。	I possessed by a spirit of abandon.
3	捜索は完全に打ち切られた。	The search was completely abandoned.
4	平然と子供を見捨てた。	He coolly abandoned his child.
⋮	⋮	⋮

【図４】

識別子	日本語	英語
1	計算、そろばん、玉、動かす、行なう	Computation , perform , abacus , manipulate , counter
2	私、どう、いい、気持ち、取りつかれる	I , possess , spirit , abandon
3	捜索、完全、打ち切られる	search , completely , abandon
4	平然、子供、見捨てる	He coolly abandon , child
⋮	⋮	⋮

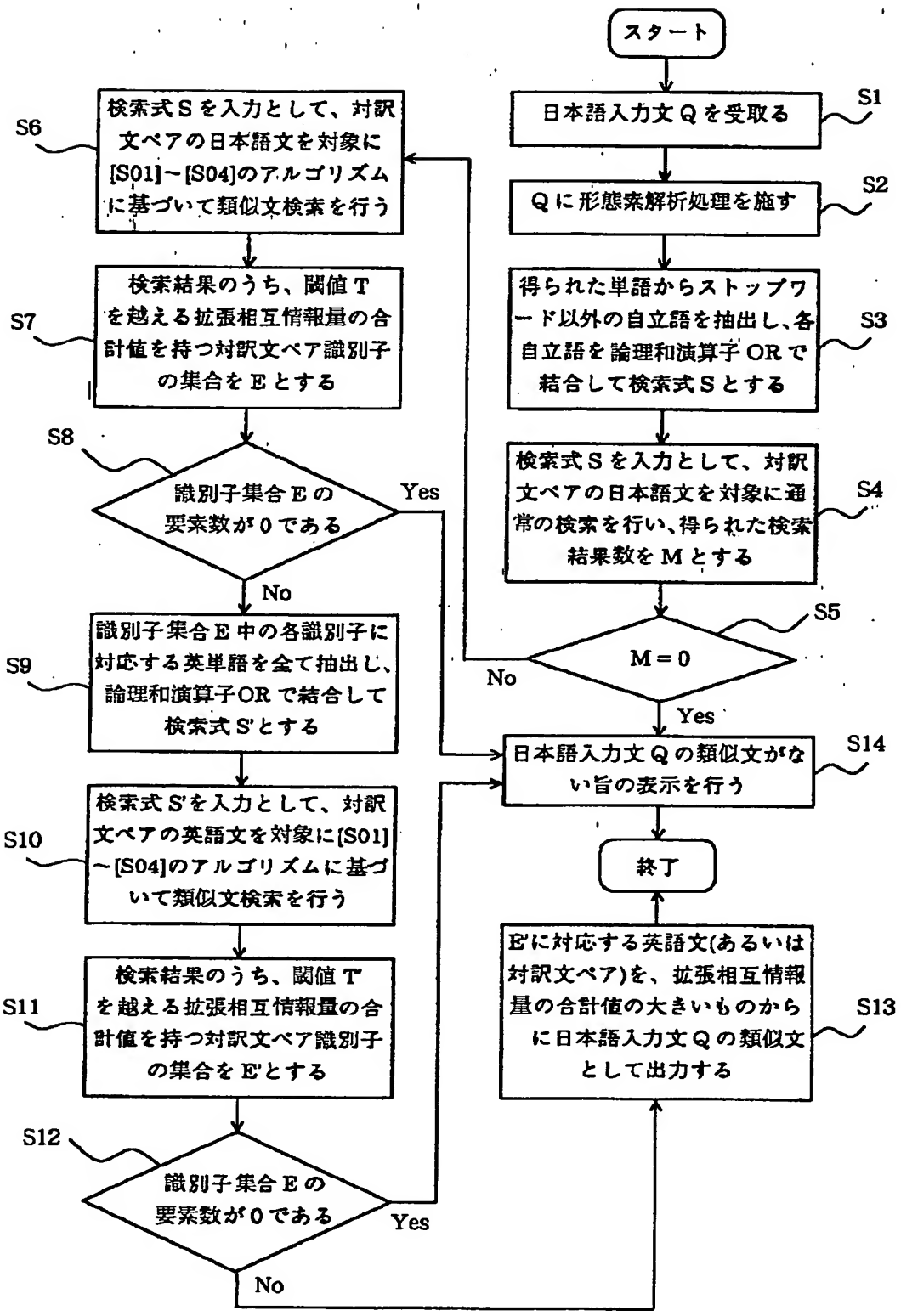
【図６】

Clouds blanketed the sky. （雲が空をおおっていた）	[9／360,000]
The sky was heavily blanketed with clouds. （空は雲で一面厚くおおわれていた）	[10／360,000]
The clouds darkened the sky. （雲が空を暗くした）	[13／360,000]
The sky darkened with thunder clouds. （空は雷雲で暗くなった）	[14／360,000]
The sky is darkened with clouds. （雲が出て空が暗くなっている）	[15／360,000]
The sky is blotted out by clouds. （空は雲で一面おおわれている）	[18／360,000]
The sky is blotted out by clouds. （雲に覆されて空がまるで見えない）	[19／360,000]

【図７】

He buried himself in anonymity. （匿名に身を隠した）	[5／360,000]
preserve [keep , maintain] one's anonymity （名を伏せておく）	[7／360,000]
She preferred to veil her identity. （本名を隠すほうを好んだ）	[8／360,000]
Name withheld by Request （希望により匿名、匿名希望）	[18／360,000]
He asked for his name to be withheld. （自分の名を公にしてくれるなと言った）	[15／360,000]

【図５】


スタート
日本語入力文Ｑを受取る
S1
Ｑに形態素解析処理を施す
S2
得られた単語からストップワード以外の自立語を抽出し、各自立語を論理和演算子ORで結合して検索式Ｓとする
S3
検索式Ｓを入力として、対訳文ペアの日本語文を対象に通常の検索を行い、得られた検索結果数をＭとする
S4
Ｍ＝0
S5
No
Yes
日本語入力文Ｑの類似文がない旨の表示を行う
S14
終了
E'に対応する英語文(あるいは対訳文ペア)を、拡張相互情報量の合計値の大きいものから日本語入力文Ｑの類似文として出力する
S13
検索式Ｓを入力として、対訳文ペアの日本語文を対象に[S01]～[S04]のアルゴリズムに基づいて類似文検索を行う
S6
検索結果のうち、閾値Ｔを越える拡張相互情報量の合計値を持つ対訳文ペア識別子の集合をＥとする
S7
識別子集合Ｅの要素数が０である
S8
Yes
No
識別子集合Ｅ中の各識別子に対応する英単語を全て抽出し、論理和演算子ORで結合して検索式Ｓ'とする
S9
検索式Ｓ'を入力として、対訳文ペアの英語文を対象に[S01]～[S04]のアルゴリズムに基づいて類似文検索を行う
S10
検索結果のうち、閾値Ｔ'を越える拡張相互情報量の合計値を持つ対訳文ペア識別子の集合をＥ'とする
S11
識別子集合Ｅの要素数が０である
S12
Yes
No

【図８】

	従来技術１	本発明
特許に抵触する。	1	2
彼は言い訳が上手だ。	0	3
目に異物が入る。	4	3
お手紙の返事が遅くなりまして申し訳ございません。	5	6
金のない乞食。	0	1
ざわめきが大きくなった。	1	3
そそっかしくて失敗した。	0	2
よぼよぼのお年寄り。	1	1
劇が開演する。	0	3
クリスマスシーズンがやって来た。	0	3
前衛的な文芸作品。	0	4
ニックネームをつける。	1	4
ホルモン剤で治療する。	0	1
雲が立ち込めてきた。	0	7
香料が発散する。	1	3
展覧会のオープニング	1	4
リサイタルでデビューする。	0	1
言論の自由を弾圧する。	1	3
哀れっぽい鳴き声。	2	4
鋭いウイットがある。	2	8
合計	20	66

【手続補正書】
【提出日】平成１１年９月２２日（１９９９．９．２２）
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】特許請求の範囲
【補正方法】変更
【補正内容】
【特許請求の範囲】
【請求項１】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索装置において、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、
第１言語で書かれた検索質問文を受け付ける検索文入力手段と、
受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、
第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、当該対訳文ペア格納手段に格納されている第２言語で書かれた訳文の集合を対象として検索する第２の検索手段と、
を有することを特徴とする対訳文検索装置。
【請求項２】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた文を検索する対訳文検索装置において、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、
第２言語で書かれた文を複数格納する文格納手段と、
第１言語で書かれた検索質問文を受け付ける検索文入力手段と、
受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、
第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、当該対訳文ペア格納手段に格納されている第２言語で書かれた訳文集合及び文格納手段に格納されている第２言語で書かれた文集合を対象として検索する第２の検索手段と、
を有することを特徴とする対訳文検索装置。
【請求項３】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた文を検索する対訳文検索装置にお

いて、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアを複数格納する対訳文ペア格納手段と、
第２言語で書かれた文を複数格納する文格納手段と、
第１言語で書かれた検索質問文を受け付ける検索文入力手段と、
受け付けた検索質問文に基づいて対訳文ペア格納手段に格納されている第１言語で書かれた文の集合を対象として検索処理する第１の検索手段と、
第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を、文格納手段に格納されている第２言語で書かれた文集合を対象として検索する第２の検索手段と、
を有することを特徴とする対訳文検索装置。

【請求項４】　請求項１乃至請求項３のいずれか１項に記載の対訳文検索装置において、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から所定の基準に基づいた重要語を抽出し、当該重要語を用いて第２言語で書かれた類似文を検索することを特徴とする対訳文検索装置。

【請求項５】　請求項１乃至請求項３のいずれか１項に記載の対訳文検索装置において、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から重要語を抽出するとともに重要語に重要度を付与し、当該重要語及び重要度を用いて第２言語で書かれた類似文を検索し、
更に、対訳文ペア格納手段に格納されている第２言語で書かれた文の集合Ａと、第１の検索手段で検索された第１言語で書かれた文に対応する第２言語で書かれた文の集合Ｂと、集合Ｂ中に出現する全単語の集合Ｃに関して、
集合Ｂに含まれる文の数である第１の値と、集合Ｂ中に出現する単語を重要語候補として各重要語候補を含む集合Ｂ中の文の数である第２の値と、各重要語候補を含む集合Ａ中の文の数である第３の値を求め、これら３種の値を変数として各重要語候補の重要度を算出し、これら重要度に基づいて重要語候補中から重要語が決定されることを特徴とする対訳文検索装置。

【請求項６】　請求項２又は請求項３に記載の対訳文検索装置において、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文から重要語を抽出するとともに重要語に重要度を付与し、当該重要語及び重要度を用いて第２言語で書かれた類似文を検索し、
更に、対訳文ペア格納手段に格納されている第２言語で書かれた文の集合と文格納手段に格納されている第２言語で書かれた文の集合の和である集合Ａと、第１の検索手段で検索された第１言語で書かれた文に対応する第２言語で書かれた文の集合Ｂと、集合Ｂ中に出現する全単語の集合Ｃに関して、
集合Ｂに含まれる文の数である第１の値と、集合Ｂ中に出現する単語を重要語候補として各重要語候補を含む集合Ｂ中の文の数である第２の値と、各重要語候補を含む集合Ａ中の文の数である第３の値を求め、これら３種の値を変数として各重要語候補の重要度を算出し、これら重要度に基づいて重要語候補中から重要語が決定されることを特徴とする対訳文検索装置。

【請求項７】　請求項５又は請求項６に記載の対訳文検索装置において、
第２の検索手段は、集合Ａ中に含まれる文書の数をＭとし、第１の値をα、重要語候補ごとの第２の値をβ、重要語候補ごとの第３の値をγとした場合に、
拡張相互情報量＝ｌｏｇ｜（Ｍβ）／（$\alpha\gamma$）｜
拡張Ｔ－ｓｃｏｒｅ＝Ｍ｜（Ｍ$\beta-\alpha\gamma$）／（$\alpha\gamma$）｜
拡張Ｄｉｃｅ－ｃｏｅｆｆｉｃｉｅｎｔ＝２β／（$\alpha+\gamma$）
のいずれかの値を各重要語候補の重要度とすることを特徴とする対訳文検索装置。

【請求項８】　請求項１乃至請求項４のいずれか１項に記載の対訳文検索装置において、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた文に基づいて、第２言語で書かれた訳文に類似する文を検索する際に、ベクトル空間モデルを用いることを特徴とする対訳文検索装置。

【請求項９】　請求項１乃至請求項８のいずれか１項に記載の対訳文検索装置において、
第１の検索手段は、受け付けた検索質問文に類似する文を検索するとともに当該類似する文に重要度を付与し、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた文及び第１の検索手段によって付与された重要度に基づいて、第２言語で書かれた訳文に類似する文を検索することを特徴とする対訳文検索装置。

【請求項１０】　請求項１乃至請求項９のいずれか１項に記載の対訳文検索装置において、
検索文入力手段は、第１言語で書かれた検索質問文を受け付けるとともに、当該検索質問文を複数の単語或いは節に分割し、
第１の検索手段は、分割された各単語或いは各節を用いて、受け付けた検索質問文に類似する第１言語で書かれた文を検索し、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応して対訳文ペア格納手段に格納されている第２言語で書かれた訳文に類似する文を検

索し、
更に、各単語或いは各節ごとに第２の検索手段により検索された第２言語で書かれた複数の文の中から、検索結果を所定の重要度を基準として選択する検索結果統合手段を有することを特徴とする対訳文検索装置。
【請求項１１】　請求項１、請求項２、請求項４、請求項５、請求項８、請求項９、請求項１０のいずれか１項に記載の対訳文検索装置において、
第２の検索手段は、第１の検索手段により検索された第１言語で書かれた文に対応する第２言語で書かれた訳文と共に、当該第１言語で書かれた文も対訳文ペア格納手段から取得し、当該第１言語で書かれた文と第２言語で書かれた訳文とからなる対訳文ペアに類似する対訳文ペアを対訳文ペア格納手段に格納されている対訳文ペアの集合から検索することを特徴とする対訳文検索装置。
【請求項１２】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、
第１言語で書かれた検索質問文を受け付ける検索文入力機能と、
メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、
第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として検索する第２の検索機能と、
をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。
【請求項１３】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、
第１言語で書かれた検索質問文を受け付ける検索文入力機能と、
メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、
第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合及び当該ペアデータとは別個にメモリに格納されている第２言語で書かれた文の集合を対象として検索する第２の検索機能と、
をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。
【請求項１４】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータに検索させるための対訳文検索プログラムを記憶した記憶媒体において、
第１言語で書かれた検索質問文を受け付ける検索文入力機能と、
メモリに記憶されている第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として検索処理する第１の検索機能と、
第１の検索機能により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータとは別個にメモリに格納されている第２言語で書かれた文の集合を対象として検索する第２の検索機能と、
をコンピュータに実現させるための対訳文検索プログラムをコンピュータにより読み出し可能に記憶したことを特徴とする記憶媒体。
【請求項１５】　請求項１２乃至請求項１４のいずれか１項に記載の対訳文検索プログラムを記憶した記憶媒体において、
記憶媒体にはペアデータが読み出し自在に記憶されており、
対訳文検索プログラムは、当該ペアデータを記憶媒体から読み出してコンピュータに備えられているメモリに格納する機能を含んでいることを特徴とする対訳文検索プログラムを記憶した記憶媒体。
【請求項１６】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータを使用して検索する対訳文検索方法において、
第１言語で書かれた検索質問文を受け付け、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索を行うことを特徴とする対訳文検索方法。
【請求項１７】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、
第１言語で書かれた検索質問文を受け付け、
記憶手段に格納されている、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを

用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索を行うことを特徴とする対訳文検索方法。

【請求項１８】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、
第１言語で書かれた検索質問文を入力し、
第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索を行い、
前記第２の検索で得られた結果を表示することを特徴とする対訳文検索方法。

【請求項１９】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文を検索する対訳文検索方法において、
第１言語で書かれた検索質問文を入力し、
記憶手段に格納されている、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索を行い、
前記第２の検索で得られた結果を表示することを特徴とする対訳文検索方法。

【請求項２０】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータを用いて検索する対訳文検索方法において、
第１言語で書かれた検索質問文を入力し、
記憶手段に格納されている、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合を対象として第２の検索を行い、
前記第２の検索で得られた結果を表示することを特徴とする対訳文検索方法。

【請求項２１】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータを用いて検索する対訳文検索方法において、
第１言語で書かれた検索質問文を入力し、
記憶手段に格納されている、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータ中の第２言語で書かれた訳文集合及び当該ペアデータとは別個に記憶手段に格納されている第２言語で書かれた文の集合を対象として第２の検索を行い、
前記第２の検索で得られた結果を表示することを特徴とする対訳文検索方法。

【請求項２２】　第１言語で書かれた検索質問文に基づいて第２言語で書かれた訳文をコンピュータを用いて検索する対訳文検索方法において、
第１言語で書かれた検索質問文を入力し、
記憶手段に格納されている、第１言語で書かれた文とそれに対応する第２言語で書かれた訳文とのペアデータを用いて、受け付けた検索質問文に基づいて第１言語で書かれた文の集合を対象として第１の検索を行い、
前記第１の検索により検索された第１言語で書かれた文に対応するペアデータ中の第２言語で書かれた訳文に類似する文を、当該ペアデータとは別個に記憶手段に格納されている第２言語で書かれた文の集合を対象として第２の検索を行い、
前記第２の検索で得られた結果を表示することを特徴とする対訳文検索方法。

フロントページの続き

(72)発明者　舘野　昌一
神奈川県足柄上郡中井町境430　グリーンテクなかい　富士ゼロックス株式会社内

Ｆターム(参考)　5B075 ND03 NK02 NK32 PP25 PR08 QM08
5B091 AA03 CA02 DA04